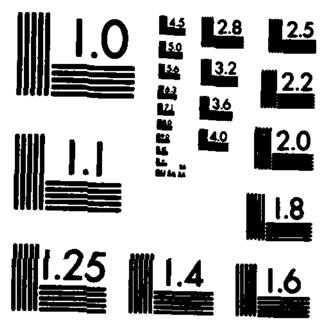
AD-A120 954 ON THE FALLACY OF THE LIKELIHOOD PRINCIPLE(U) WISCONSIN 1/1
UNIV-MADISON MATHEMATICS RESEARCH CENTER H AKAIKE
JUN 82 MRC-TSR-2395 DAAG29-80-C-0041

UNCLASSIFIED F/G 12/1 NL

END
DATE
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MRC Technical Summary Report #2395

ON THE FALLACY OF THE LIKELIHOOD PRINCIPLE

Hirotugu Akaike

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

June 1982

(Received May 17, 1982)

Approved for public release
Distribution unlimited

Sponsored by:

U. S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

ON THE FALLACY OF THE LIKELIHOOD PRINCIPLE

Hirotugu Akaike*

Technical Summary Report #2395
June 1982

ABSTRACT

By using the direct and inverse binomial experiments it is shown that there is a situation where Birnbaum's basic axiom of mathematical equivalence and the likelihood principle is a tautology. This observation disqualifies the Birnbaum's proof of the likelihood principle based on the axioms of mathematical equivalence and conditionality. The implication of this disproof of Birnbaum's argument for Bayesian statistics is briefly discussed.

AMS (MOS) Subject Classifications: 62A10, 62A15.

Key Words: Likelihood Principle; Conditionality; Sufficiency; Bayesian Statistics; Prior Distribution.

Work Unit Number 4 - Statistics and Probability

*The Institute of Statistical Mathematics, Tokyo, Japan.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

The likelihood principle demands that same inference should be made from data with identical likelihood function. It has often been considered that a proof of the principle was given by Allan Birnbaum and that the principle holds automatically in Bayesian statistics. In this paper, using a simple example of direct and inverse binomial experiments, it is shown that Birnbaum's proof reduces to a tautology. It is expected that this semantic disproof of Birnbaum's result will eliminate the confusion about the relation between the likelihood principle and Bayesian statistics and will eventually contribute to the development of the latter.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

ON THE FALLACY OF THE LIKELIHOOD PRINCIPLE

Hirotugu Akaike*

1. INTRODUCTION AND SUMMARY

There has been a continuous discussion of the likelihood principle which demands that the statistical inferences based on the data with identical likelihood functions should be indentical; see, for example Birnbaum (1962 1972) and references therein.

The likelihood principle has been considered to be providing a strong support for Bayesian statistics (de Finetti 1972, Lindley 1972, Savage 1962). In particular, the direct and inverse binomial experiments with unknown probability of occurrence is often quoted as a typical example where the conventional and Bayesian approaches differ significantly.

The likelihood principle gained much support by its "proof" given by Birnbaum (1962) who derived the principle from the principles of conditionality and sufficiency. The "proof" was later improved by replacing the sufficiency principle by a logically weaker axiom, or piciple, of mathematical quivalence (Birnbaum 1972). The force of Birnbaum's "proof" lies in the fact that the two axioms of conditionality and mathematical equivalence are undeniably "obvious" or "natural". The definitions of these axioms are given in the next section.

Many authors have discussed the likelihood principle. A good summary of the interplay between various piciple of inference related with the

*The Institute of Statistical Mathematics, Tokyo, Japan.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041.

likelihood principle is given by Dawid (1977). Particularly notable in these works is the one by Durbin (1970) who developed a strong argument against Birnbaums' simultaneous application of the conditionality and sufficiency principles. Apparently Birnbaum's introduction of the axiom of mathematical equivalence was intended to replace the axiom of sufficiency by one that is more "obvious" and thus will be more independent of the conditionality principle.

The purpose of the present paper is to show, by using the example of direct and inverse binomial experiments, that the axiom of mathematical equivalence is by no means "obvious" as claimed by Birnbaum. In this example, under the assumption of the axiom of conditionality, the relation between the axiom of mathematical equivalence and that of likelihood reduces to a tautology, which shows that the difficulty pointed out by Durbin still persists.

The relation between the likelihood principle and Bayesian statistics is discussed briefly at the end of this paper.

2. BIRNBAUM'S AXIOMS AND THE LIKELIHOOD PRINCIPLE

Following Birnbaum (1972) an experiment E is defined by $E = (\Omega, S, f)$, where $S = \{x\}$ is the discrete sample space, $\Omega = \{\theta\}$ is the parameter space and $f = f(x,\theta) = \text{Prob}(X = x|\theta)$. For each x, (E,x) is used to represent (an instance of) statistical evidence. For E and E^* with common parameter space the judgement of the equivalence of the statistical evidences (E,x) and (E^*,x^*) is represented by $E_v(E,x) = E_v(E^*,x^*)$.

A statistic $h(x)$ is called ancillary if it satisfy the relation

$$f(x,\theta) = g(h)f(x|h,\theta)$$

where $g(h) = \text{Prob}\ \{(x; h(x) = h)\}$ is independent of θ. For each possible

value h of an ancillary statistic $h(x)$ E_h is defined by $E_h = (\Omega, S_h, f_h)$ where $S_h = \{x;\ h(x) = h\}$ and $f_h = f(x|h,\theta)$. Then E may be considered as a mixture experiment composed of components E_h having respective probabilities $g(h)$.

Birnbaum's axioms required for the recent discussion are:

Conditionality (C): $Ev(E,x) = Ev(E_h,x)$ where h is an ancillary statistic and $h = h(x)$.

Mathematical equivalence (M): If $f(x,\theta) = f(x',\theta)$ for all $\theta \in \Omega$, then $Ev(E,x) = Ev(E,x')$.

Likelihood (L): If for some $c > 0$, $f(x,\theta) = c\, f^*(x^*,\theta)$ for all $\theta \in \Omega$, then $Ev(E,x) = Ev(E^*,x^*)$.

The likelihood principle dictates that the evidence produced by a valid statistical inference procedure must satisfy (L). A support for the principle comes from Birnbaum's proof that (C) and (M) jointly imply (L). This is often considered a "proof" of the likelihood principle.

Birnbaum also showed that (L) implies (C) and (M). Thus, under the assumption of (C), (M) and (L) are mathematically equivalent.

3. THE FALLACY

The force of support for the likelihood principle rendered by Birnbaum's "proof" is based on the observation that (C) and (M) are "natural" or "obvious". When (C) is accepted as one of the premises then, due to the

logical equivalence of (M) and (L), the force of the support depends entirely on the "obviousness" of (M).

Birnbaum adopted (M) simply by assuming that it is the simplest formalization of the equivalent statistical evidence. This argument is obviously extra-mathematical and is concerned with the meaning of (M). We will show that this extra-mathematical argument is unfounded.

Consider the direct binomial experiment $E = (\Omega, S, f)$ defined by $f(x,\theta) = {}_nC_x\, \theta^x(1-\theta)^{n-x}$, $S = \{0,1,\ldots,n\}$ and $\Omega = \{\theta;\ 0 \leq \theta \leq 1\}$. Also consider the inverse binomial experiment $E' = (\Omega, S', f')$ defined by $f'(y,\theta) = {}_{y-1}C_{m-1}\, \theta^m(1-\theta)^{y-m}$, $S' = \{m, m+1, \ldots\}$ and with the same Ω as that of E. We assume that n and m are positive integers such that $n > m$.

For $x = m$ and $y = n$ we have $f(m,\theta) = cf'(n,\theta)$ $(c > 0)$. Following Birnbaum we generate the mixture experiment $E^* = (\Omega, S^*, f^*)$, where $\Omega = \{\theta\}$, $S^* = \{(u,v);\ u,v = 0,1,2,\ldots\}$ and f^* is defined by

$$f^*((u,v),\theta) = \begin{cases} hf(x,\theta) & \text{for } (u,v) = (n,x) \\ (1-h)f'(y,\theta) & \text{for } (u,v) = (m,y) \\ 0 & \text{otherwise,} \end{cases}$$

where $h = 1/(1+c)$. For this experiment we have $f^*((n,m),\theta) = hf(m,\theta) = (1-h)f'(n,\theta) = f^*((m,n),\theta)$. In this case (M) says that $Ev(E^*,(n,m)) = Ev(E^*,(m,n))$.

Birnbaum's justification of (M) rests on the argument that by relabelling (n,m) by (m,n), and vice versa, no observable change is introduced into the probabilistic structure of the experiment; see the discussion of the example at the beginning of p.889 of Birnbaum (1972). However, in the present example, by observing (n,m) we know that the data were generated by the direct binomial experiment. Also, by observing (m,n) we know that it came from the inverse binomial experiment. The empirical

meaning of the relabelling is to make a false report (n,m) when actually (m,n) is observed. Whether such reporting is acceptable or not is not an obvious matter. It is acceptable only when we are willing to ignore the difference between (m,n) and (n,m).

When (C) is invoked the acceptance of the false reporting procedure is equivalent to accepting the equality $E_v(E,m) = E_v(E',n)$, i.e., the likelihood axiom. Thus, under the assumption of (C), the relabelling assumed by (M) is neither more nor less compelling than the likelihood axiom (L) itself. In this situation the relation between (M) and (L) is a tautology. A tautology is logically a truth. However, it does not provide any new information that adds to our knowledge. This observation constitutes a semantic disproof of Birnbaum's "proof" of the likelihood principle.

4. DISCUSSION

Birnbaum's "proof" has produced significant effect on Bayesian statisticians. Savage's reaction to Birnbaum's "proof" (Savage 1962) typically represents the interpretation by a Bayesian statistician of the relation between the likelihood principle and Bayesian statistics. Lindley (1972) changed his interpretation of the principle from that stated in the unquestionable definition given in Lindley (1965), where the equality of prior distributions was an explicit prerequisit of the principle, to that of Savage who simply assumed the uniqueness of the prior distribution. Apparently Lindley (1972) accepted Birnbaum's "proof" as a confirmation of the principle.

It is an unlucky coincidence that the "proof" strengthened the then-growing misconception that the principle automatically holds in Bayesian statistics. Many text books adopt this view. Such a view is obviously in contradiction with the original position of the subjective theory of

probability which tells that every prior information should be taken into account when developing a probability distribution of an unknown. It is rather surprising to see that even the typical subjectivist de Finetti violates this basic position by simply admitting the likelihood principle (de Finetti 1972). Such a breach of the basic teaching has unfailingly lead Bayesian statistics to a stalemate.

Strongly negative reactions have been shown against the present author's discussion of the necessity of including the sampling scheme into prior information when developing a prior distributions for direct or inverse binomial experiments (Akaike 1980). This looks natural when we recognize that Birnbaum's argument attracted even philosophers' interest (Kyburg 1974, Seidenfeld 1979). However, scientifically minded users of Bayesian statistics do not show any doubt in taking into account the property of sampling scheme in developing a prior distribution. A typical example is Jeffreys (1946) who defines an ignorance prior distribution by using the information supplied by the specification of the likelihood function. Box and Tiao (1973) convincingly discuss the rationality of using different ignorance priors for direct and inverse binomial experiments.

It seems fair to say that the likelihood principle is unconditionally supported only by those Bayesian statisticians who tend to see some uniquely defined objective meaning in the parameter or its probability distribution. An argument for the use of such an "objective" prior distribution as our "subjective" prior distribution, when it does exist, is developed in Akaike (1978). However, when we do not have such a strong prior information, we cannot deny the possibility of using the information about the expected behavior of the posterior distribution in choosing a prior distribution. Such a procedure of selection naturally induces the dependence of the prior

distribution on the sampling scheme.

The semantic disproof of Birnbaum's derivation of the likelihood principle presented in this paper will emancipate statisticians from the spell of the yet unfounded likelihood principle. Statisticians are still free in developing their prior distributions for each particular problem. They may use the information on the assumed sampling scheme in choosing a prior distribution. This freedom will give Bayesian statisticians more chance to appreciate the common sense embodied in conventional statistics and thus eventually enhance the vitality of Bayesian statistics.

ACKNOWLEDGEMENT

The author is grateful to G.E.P. Box, A.P. Dawid, T. Leonard and C.F. Wu for helpful discussions and comments.

REFERENCES

Akaike, Hirotugu (1974), "An Objective Use of Bayesian Models", Annals of the Institute of Statistical Mathematics, 29, 711-720.

Akaike, Hirotugu (1980), "Likelihood and The Bayes Procedure, with discussion", in Bayesian Statistics, eds. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, Valencia, Spain: University Press, 143-166 and 185-203.

Birnbaum, Allan (1962), "On The Foundation of Statistical Inference", with discussion", Journal of the American Statistical Association, 57, 269-326.

Birnbaum, Allan (1972), "More on Concepts of Statistical Evidence", Journal of American Statistical Association, 67, 858-861.

Box, George E. P. and Tiao, George C. (1973), Bayesian Inference in Statistical Analysis, Reading, Massachusetts: Addison-Wesley.

Dawid, A. P. (1977), "Conformity of Inference Patterns", in Recent Developments in Statistics, eds. J. R. Barra, B. van Cutsem, F. Brodeau and G. Romier, Amsterdam: North-Holland, 245-256.

de Finetti, Bruno (1972), Probability, Induction and Statistics, New York: Wiley.

Durbin, J. (1970), "On Birnbaum's Theorem on the Relation Between Sufficiency, Conditionality and Likelihood", Journal of the American Statistical Association, 65, 395-398.

Jeffreys, H. (1946), "An Invariant Form for the Prior Probability in Estimation Problems", Proceedings of the Royal Society of London, Ser. A 186, 453-461.

Kyburg, Henry E. (1974), The Logical Foundation of Statistical Inference, Dordrecht-Holland: Reidel.

Lindley, Dennis V. (1965), Introduction to Probability and Statistics, PART 2, Cambridge University Press.

Lindley, Dennis V. (1972), Bayesian Statistics, A Review, Philadelphia: Society for Industrial and Applied Mathematics.

Savage, L.J. (1962), "Discussion on Birnbaum's Paper", Journal of the American Statistical Association, 57, 307-308.

Seidenfeld, Teddy (1979), Philosophical Problems of Statistical Inference, Dordrecht-Holland: Reidel.

[illegible]/jh

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 2395	2. GOVT ACCESSION NO.: AD-A120 954
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): On the Fallacy of the Likelihood Principle	5. TYPE OF REPORT & PERIOD COVERED: Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Hirotugu Akaike	8. CONTRACT OR GRANT NUMBER(s): DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 4, Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, P.O. Box 12211, Research Triangle Park, North Carolina 27709	12. REPORT DATE: June 1982
	13. NUMBER OF PAGES: 10
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

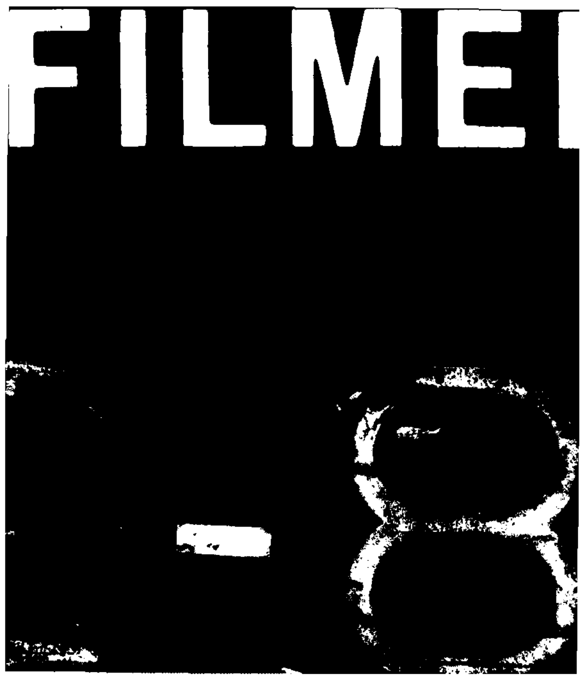
18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Likelihood Principle, Conditionality, Sufficiency, Bayesian Statistics, Prior Distribution.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

By using the direct and inverse binomial experiments it is shown that there is a situation where Birnbaum's basic axiom of mathematical equivalence and the likelihood principle is a tautology. This observation disqualifies the Birnbaum's proof of the likelihood principle based on the axioms of mathematical equivalence and conditionality. The implication of this disproof of Birnbaum's argument for Bayesian statistics is briefly discussed.


FILME
-8